

What is genomics?

THE HUMAN BODY consists of trillions of cells. Almost all contain an entire **genome**—the complete set of inherited genetic information encoded in our DNA. When humans reproduce, the parents’ sperm and egg DNA combine to contribute a genome’s worth of genetic information to the fertilized embryo. That same information is in each of the cells that eventually make up an organism.

Some segments of DNA, called **genes** or “coding” DNA, contain the chemical recipe that determines particular traits; **genetics** is the study of the inheritance and function of these genes. Scientists now estimate that humans have about 30,000 genes, located along threadlike, tightly coiled strands of DNA called **chromosomes**. Genes, however, are only about three percent of human DNA; the rest is “noncoding” DNA. Within these noncoding regions of the



genome is the information that determines when and where genes are active—for example, in which cell types and at what stages in the life of an organism. **Genomics** is the study of the entire set of DNA sequences—both coding and noncoding DNA.

Over the past decade, the decoding of the genomes of human beings and other important organisms has sparked an extraordinary biological revolution. The information and technology of genomics is transforming our understanding of

human evolution, the mechanisms of disease, the relationship between heredity and environment, and our ancient connection with all forms of life. In the next few years we will see many exciting discoveries leading to a better understanding of the complexity of life, as well as new drugs, vaccines, and diagnostics and less expensive, more efficient, and safer ways to restore the environment.

DNA: life’s code

The double-stranded DNA (deoxyribonucleic acid) molecule contains the four basic chemical units of life’s code: the **nucleotide bases** adenine (**A**), guanine (**G**), cytosine (**C**), and thymine (**T**). These combine into the base pairs **AT**, **TA**, **GC**, and **CG**. The paired bases form the “rungs” of a structure that looks like a twisted rope ladder—the famous double helix. Sugar and phosphate molecules form the outer edges.

Translating the code: DNA → RNA → proteins

Following the DNA recipe, our cells manufacture the **proteins** that are responsible for the structure and functioning of our bodies. Proteins are involved in many of the body’s life processes, including growth, repair, digestion, and aging. Many proteins are **enzymes** that can trigger or accelerate chemical reactions. Others are **transporters**, such as hemoglobin, found in red blood cells, which takes oxygen from the lungs to the body’s cells.

Proteins are produced from the DNA recipe in two basic steps:

Transcription

Because DNA never leaves the cell’s nucleus, a “messenger” must be created to move its information out into the cell. First, a key enzyme called **RNA polymerase** makes the DNA unwind and “unzip” by breaking the hydrogen bonds between the bases in the paired strands (1). This process forms two complementary strands—a **coding strand** (2) and a **template strand** (3).

**RNA** (ribonucleic acid), a single-stranded molecule very similar to DNA, is then created as nucleotide bases are strung together by the RNA polymerase in a sequence determined by the DNA template (4). The new RNA strand has the same information as the original coding strand, with one exception: **U** (uracil) nucleotides substitute for the **T**s in the DNA. The resulting strand is called **messenger RNA** (mRNA).

Next, the mRNA travels out into the body of the cell—the **cytoplasm**—and attaches to a **ribosome**, the cell’s protein factory. Every cell has thousands of these tiny factories.

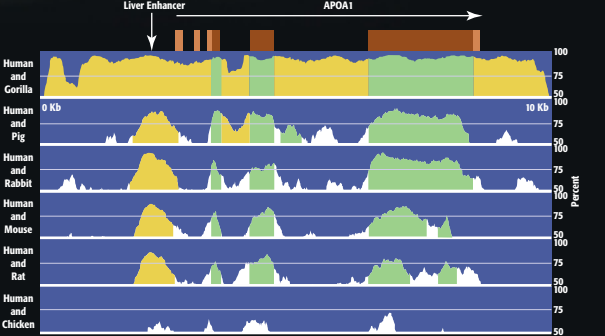
Translation

The **As**, **Cs**, **Gs**, and **Us** in the mRNA are read by the ribosome as three-letter “words,” called **codons**, which are known as the **genetic code**:  
**AUG GAA UUC UCG CUC**  
There are 64 codons, representing the 20 or so **amino acids** that are the building blocks of proteins. These numbers are not equal because sometimes two or more codons code for the same amino acid. There are also “start” and “stop” codons that determine where the protein chain begins and ends.

Now the ribosome has enough information to manufacture, or **synthesize**, a protein. The ribosome moves along the mRNA strand and reads its sequence, one codon at a time (5). With the help of another type of RNA called **transfer RNA** (tRNA), the ribosome adds amino acids one by one to the growing chain, called a **polypeptide chain** (6). When it’s complete, the chain folds into a specific shape dictated by the amino acid sequence (7) and becomes a protein; its shape determines the protein’s function in the body. The translation process from DNA to protein is complete.

Understanding DNA

Comparing the DNA sequence patterns of humans side-by-side with those of well-studied “model organisms” such as the fruit fly, mouse, pufferfish, and sea squirt is one of the most powerful strategies for identifying human genes and determining how they’re regulated and what they do. **Conserved** sequences—DNA patterns that we share with other organisms—are likely to have important functions or they would have disappeared as the organisms evolved.



This computer plot shows how similar a segment of DNA is in the genomes of the human, gorilla, pig, rabbit, mouse, rat and chicken. Visualizations like these make it easier for scientists to identify conserved regions of DNA that could be important in regulating gene and protein function (green areas indicate conserved coding gene segments; yellow areas are conserved noncoding elements).

Along with helping identify genes and their functions, comparative genomics is shedding light on the functions of the noncoding sequences of DNA found within and between the genes. These segments can regulate gene **expression**, the process involved in determining when and where in the organism a given gene is turned on or off. Understanding the complex orchestration of gene and protein networks is a crucial aspect of contemporary biomedical research.

The web of life

We may not look alike, but humans, cows, fish, and microbes have a lot in common. We all retain similar DNA sequences inherited from our shared ancestors who lived hundreds of millions of years ago. Humans share many genes with mice, frogs, flies, and even bacteria and yeast, since their essential functions have been inherited intact over long periods of time.

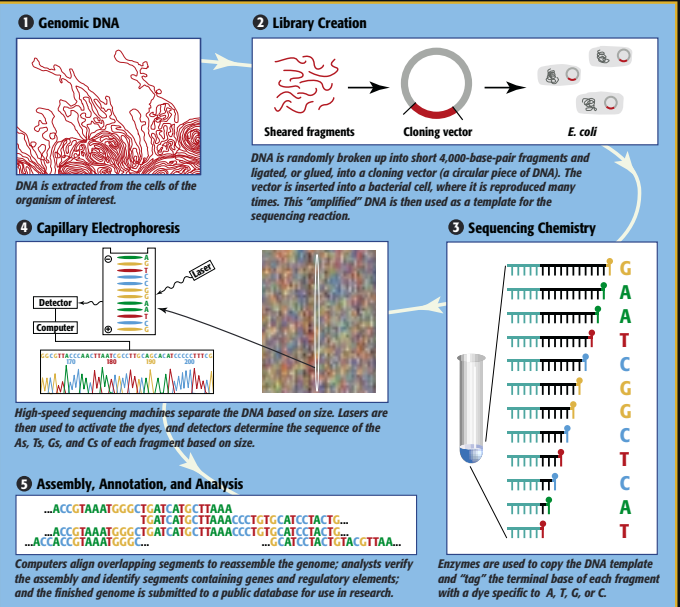


Species	Chromosomes	Genes	Base Pairs
<b>Human</b> ( <i>Homo sapiens</i> )	<b>46 (23 pairs)</b>	<b>28-35,000</b>	<b>~3.1 billion</b>
<b>Mouse</b> ( <i>Mus musculus</i> )	<b>40</b>	<b>22.5-30,000</b>	<b>~2.7 billion</b>
<b>Pufferfish</b> ( <i>Fugu rubripes</i> )	<b>44</b>	<b>~31,000</b>	<b>~365 million</b>
<b>Malaria Mosquito</b> ( <i>Anopheles gambiae</i> )	<b>6</b>	<b>~14,000</b>	<b>~289 million</b>
<b>Sea Squirt</b> ( <i>Ciona intestinalis</i> )	<b>28</b>	<b>~16,000</b>	<b>~160 million</b>
<b>Fruit Fly</b> ( <i>Drosophila melanogaster</i> )	<b>8</b>	<b>~14,000</b>	<b>~137 million</b>
<b>Roundworm</b> ( <i>C. elegans</i> )	<b>12</b>	<b>19,000</b>	<b>~97 million</b>
<b>Bacterium</b> ( <i>E. coli</i> )	<b>1*</b>	<b>~5,000</b>	<b>~4.1 million</b>

\*Bacterial chromosomes are *chromosomes*, not true chromosomes.  
The original estimate of more than 100,000 human genes was adjusted to between 28,000 and 35,000 when the draft human genome sequence was published in February 2001.

Sequencing genomes

While the number of chromosomes, genes, and base pairs in the genomes of different organisms vary, their fundamental structures are very similar, and the techniques for sequencing and studying them are the same. Whole genome shotgun sequencing is used to determine the order of the bases of an entire genome.



Field test version (March 2003). To comment e-mail: [www@cuba.jgi-psf.org](mailto:www@cuba.jgi-psf.org) or call (925) 296-5808.



Genomics applications

Fighting disease

Some human diseases and defects are directly or indirectly caused by genetic abnormalities. Sickle cell anemia, for example, is caused by a change in just one nucleotide out of six billion. Specific genes have been associated with breast cancer, deafness, and blindness. Some illnesses are caused by complex, interacting environmental and genetic factors and cannot be explained by classical inheritance patterns. Genome studies help medical researchers understand the molecular details of these diseases so they can pursue innovative drug treatments and more quickly identify high-risk individuals who could benefit from early medical intervention. And the analysis of the genomes of disease-causing microbes, viruses, and insects, such as the human malaria parasite and its carrier, the *Anopheles* mosquito, are helping in the development of new prevention and treatment strategies.



Most cystic fibrosis is from one DNA mutation—deletion of just 3 nucleotides—causing buildup of large amounts of mucus in the lungs.

Protecting plant life

Fungi and other plant pathogens cause billions of dollars in damage every year to agricultural crops, plants, and trees. Sequencing their genomes is helping botanists and foresters find effective treatments. Better understanding of plant genetics is also improving crop yields and enhancing the nutritional value of food.



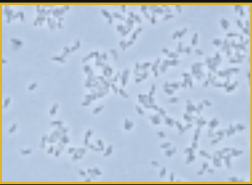
An oak tree damaged by the sudden oak death pathogen *Phytophthora ramorum*

The Human Genome...By the Numbers

- 75-100 trillion...Cells in the human body
- 3.1 billion...Base pairs in each cell
- 2.4 million...Base pairs in the largest human gene (dystrophin)
- 28,000-35,000...Genes in the human genome
- 46...Chromosomes in each cell

Harnessing nature’s technology

Microbes—nature’s simplest and most abundant organisms—can thrive under extreme conditions of heat, cold, pressure, and even radiation. By studying their genomes, scientists hope to find ways to use bacteria and other microorganisms to solve a variety of environmental problems, develop new energy sources, and improve industrial processes. Some microbes can help clean up hazardous waste sites by absorbing, transforming, or breaking down contaminants—a technique called **bioremediation**. Others can help combat global warming by absorbing, or **sequestering**, carbon from the atmosphere. And microbes can convert a wide range of organic and inorganic materials into renewable energy.

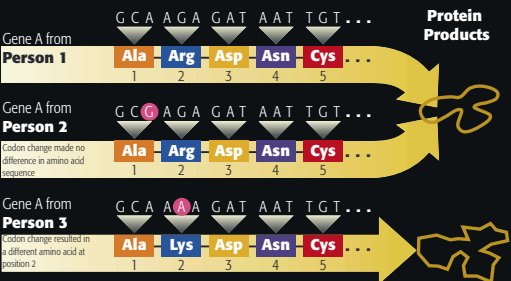


The bacterium *Rhodospseudomonas palustris* can degrade complex aromatic hydrocarbons, assimilate carbon, and provide insights into biomass and biofuel production, particularly hydrogen.

Photo: Caroline Harwood, University of Iowa

Human differences and mutations

The DNA Sequence in every human is 99.9 percent identical to that of every other human. The slight variations in our genomes are called **single nucleotide polymorphisms**, or SNPs. Scientists estimate that there are about 1.4 million locations on the genome where SNPs occur in humans. It is these small variations that contribute to individual differences. SNPs and other mutations can be caused by copying errors as DNA is reproduced, or triggered by radiation, viruses, or toxic substances in the environment. Many SNPs have no effect on cell function, but others can cause or predispose a person to disease or influence response to a drug.



DNA sequence variation in a gene can change the protein produced by the genetic code.